Digging deeper: Big data, elites and investigative research
David Miller and William Dinan
University of Bath, University of Stirling

Introduction

Digitalisation, big data and new digital research methods may pose significant risks for empirical sociology. But they also present opportunities. This chapter argues that embracing big data, exploring its uses and developing methods, theories (and ethics) to handle and conceptualise what it tells us are urgent tasks. Market research and private sector access to various forms of transactional and digital data (Savage & Burrows 2007) do offer a timely reminder of the need for a dynamic and reflexive discipline like sociology to rethink its place in the rapidly evolving social research eco-system. We would like to start our discussion by citing the closing challenge issued by McKie and Ryan in the introduction to an earlier collection (2012: 6): 'Is it our role as sociologists to describe and understand or to dig deeper and identify the social causes underpinning such complex social processes?' In our view it must be both. Obviously sociologists differ in their interests, temperament, social formation and cultural location and no sociologist can hope to do everything. But should we not aim as a discipline to produce work that is both capable of description and of tackling the big holistic, causal, contextual issues that founded the subject?

We emphasise the injunction to 'dig deeper' as in our view too little sociology digs anywhere like deep enough into social problems, far less social causes. There is far too little 'investigative' research as compared with the 'collaborative' research that seemingly every university ethics policy is developed to handle. By investigative research we have in mind a mode of inquiry that takes as its object the institutions and actors that shape social and political life, and is properly sceptical about their accounts of the world. As such the social scientist adopting an investigative approach is akin to an investigative journalist, but using the arsenal of research methods and practices of the social sciences to produce what Molotch termed 'deep journalism' (1994). We think this is particularly apt and timely as there are signs of a revival of investigative journalism embracing big data. The investigative style we are advocating here is interdisciplinary, and sociology can engage with those inside and outwith the academy who are also interested in researching the powerful using digital tools and big data. We take this notion of 'studying up' (Nader 1972; Williams 1989) as a jumping off point in sketching one means for empirical sociology to forge new forms of elite studies using big data and ICTs, exploiting the proliferation of new data sources, be they leaked, scraped or officially published through transparency, disclosure or open data projects.

Let us say a little on what we mean by 'big data', and 'digital research methods'. There is no universally agreed definition of big data, though it is conventionally used to refer to particular trends and characteristics, including the volume, variety and velocity (Stapleton 2011) of data production and sources, and how such data can be mined, manipulated and managed. In practice big data often refers to the 'digital research methods' used to collect and analyse digital data, including in particular the 'capacity to search, aggregate, and cross reference, large data sets' (boyd and Crawford 2012: 663). We use this more capacious definition since it captures more of what is done in practice and it includes the methods used to generate digital data, including, data scraping and mining as well as computer assisted data analysis. Most importantly, it refers specifically to the comparative and integrated use of different datasets. Interoperability is key both to understanding big data and its methods as well as the opportunities that - defined this way – it offers sociology.

This chapter examines the implications of publicly available data for digging deeper and developing new methods capable of offering an enhanced role for sociology. We do this by first outlining the key challenges and emerging opportunities that accompany big data. The discussion here reflects some of the concerns noted by Savage and Burrows (2007; 2009), not least the fate of established research methods and uncertainties over access to, and ownership of, data. In our view the portrait of an isolated discipline on the defensive is too pessimistic. We take up the question of how sociology should relate to the proliferation of social data by recognising the trends towards interdisciplinary work, and point to potential collaborations between computing, social science and communications studies within the academy, and with data journalists, investigative researchers and watchdog groups in civil society. This orientation exploits the reality that not all big data is private, and that there are possibilities to conduct elite studies using publicly available sources, in ways that circumvent issues of access. Critical (or investigative) scholarship on elites inevitably foregrounds questions of power, and we consider power dynamics and some of the associated ethical and political issues raised in studying contemporary power elites and power structures using digital methods and big data throughout this chapter. This discussion suggests one way out of the various crises facing sociology is to engage with debates about public sociology and equally engage in public debates about privacy, secrecy, security, accountability and trust that will inevitably require our critical understanding of the uses and abuses of big data.

The coming crisis? Empirical social science grapples with private *and* open data
Savage and Burrows (2007:887), in their widely cited contribution, suggest that sociologists are left behind by corporations able to access data 'that did not require a special effort to collect, but was the digital by-product of the routine operations of a large capitalist institution'. The ease of access is compounded by it being 'private data to which most academics have no access' (2007:887). In addition, corporations are said to have the ability to merge 'myriad sources of commercial social transactional data… without having to consider many of the ethical constraints which condition the work of academic researchers.' Some of the key challenges facing empirical sociology associated with the new data environment have been linked with questions of access, ownership and resources. In short the concern is that private entities in particular (but the point equally applies to data owned by state agencies) are relentlessly gathering information about the public, at individual and aggregate levels, that allows a granular and almost real-time knowledge of public preferences and behaviours, which are simply beyond the reach of social scientists. Sociology is denied access to this data on grounds of commercial confidentiality or national security. The worry is therefore that social scientists in the academy simply don't have a monopoly on data-gathering or useful data, and associated with this, professional social science loses some of its jurisdictional (Burrows, this volume) or privileged claims to authoritative and useful knowledge.

Without underestimating the difficulties caused by the enclosure of digital data, it is also true that big datasets are now available – or can be scraped directly – from a bewildering and burgeoning variety of sources. Not all big data is private, or stored behind firewalls on corporate or government servers, The 'open data' movement and the intrinsic nature of technological developments open up vast swathes of data to the public gaze – and potentially sociological analysis. The examples of Wikileaks and Edward Snowden illustrate the unpredictable and changing uses and status of such data. In addition, the ability to integrate more than one dataset is not confined to the corporations. Certainly, it will be necessary for the Research Councils to invest seriously in infrastructural capacity to enable the social sciences to properly take advantage of the new possibilities, but much can be done with relatively modest research support.

Savage and Burrows also suggest that key agents in the research apparatus of contemporary capitalist organizations 'now simply don't need the empirical expertise of quantitative social scientists' (2007:890-91) in particular because they have 'more effective tools than sample surveys' - the transactional data gathered by supermarkets, telecoms and internet/social media companies, and online retailers. Of course it is also the case that such data is collected or accessed by other agencies – notably those of the state and in particular the state security apparatus. Data from loyalty and credit card, mobile phones, CCTV and myriad other sources can be accessed in real time not simply in monitoring particular suspects but in assessing 'risk' levels in a particular population or locale. This could present similar kinds of challenges to sociology as discussed by Savage and Burrows.

On the other hand, the assumption that sociologists should be, or should want to be, called upon to help either corporate or state security institutions, seems not to be wholly unproblematic (Miller and Mills 2010; Miller 2013). In part it depends on varying views of what sociology is for. From the perspective of administrative research, for example, being cut off from research funding from global corporations might seem more important than from the perspective of a public sociology working with and in civil society (Burawoy 2005). In any case social scientists should analyse how these techniques are used by the state and the corporate sector and indeed the role of expertise (including perhaps sociological expertise) in constituting and reproducing the resulting social formations. In other words such developments also present challenges to Sociology in how to adequately respond in analyzing state surveillance and security or indeed the effects of digitalization on corporate power. Social science is not exactly likely to find itself in a contest for the space to critically analyse such practices.

In practice the evidence that sociologists are not needed or wanted by the powerful, is at best equivocal. Obviously this varies between industries or state agencies and in relation to various methodologies. Some state and corporate agencies seem very interested in what sociologists (and other scholars) think or can do for them. The historical trajectory here shows some similarity with that painted by Savage and Burrows in the sense that the digital era brings new research problems as well as new data and new methods for capturing and analyzing them. However, it is obvious that other dynamics interact with these processes. Thus climate change and the 'war on terror' are both issues that have come to the fore co-temporally with digitization. In the period since the late 1960s the relationship between the state, in particular and sociology has of course had its ups and downs. The period when the Conservative government renamed the 'Social Science Research Council' as the 'Economic and Social Research Council' was seen at the time as part of an attack on social science in general and indeed on sociology in particular.

The relations between the state security apparatus in the US and the UK and social science have also been seen as problematic. In particular, in the 1960s, it was precisely the (covert) interest shown by the state security apparatus in research on counterinsurgency and radicalism, which was seen, by at least some sociologists, as a problem. But that perspective was not a universal one. We now know that universities in the US collaborated with the intelligence agencies for some considerable period in the 1940s and 1950s (Diamond 1992) and that much sociological work was sponsored by (and indeed the discipline of communication studies partially created by) the psychological warfare priorities of the US state. (Simpson 1994) The exposure in 1965 of Project Camelot, in which the US aimed to 'predict and influence' social change in the developing world, was one of a number of factors

making the relationship between the secret state and sociology more distant - for a generation. (Herman 1998: 101) Following the attacks on 11 September 2001 and the launch of the 'War on Terror', state agencies in the US and UK became exceptionally interested in harnessing social science research capacity. (Miller and Mills 2010) This has of course led to a number of controversies about whether the Research Councils were too close to the security state including the occasion where the ESRC withdrew an initiative for which MI5 had provided funds. (Miller et al 2011) As we write, the case of the Liverpool University ESRC project on dissident Republicanism in Ireland has been reported as threatened by over-close relations with the security state. (Hayes and McIntyre 2014)

Studying Up with Big Data: mapping power structures and networks
In our view it is important for social scientists to investigate the activities, culture and milieu of powerful institutions and individuals in addition to studying 'down'. Arguably the quantitative method that big data really brings into its own is social network analysis (see also Ruppert in this volume on how social network analysis enacts social worlds). When it comes to interviews, the digital age again offers a myriad of new or partially new ways to gather data, including the wide-ranging tool-box of digital, documentary and ethnographic research techniques that can be put to investigative purposes. Such multi method approaches can complement and in some cases surpass the data gathering potential of interviews.

We can recall Domhoff's (2005) well-known formulation, 'there are two general aspects of any study that attempts to delineate a power structure and understand its workings'. The first is a network analysis, which 'can reveal power structures to be more or less tightly knit, more or less focused in a few organizations or individuals, and more or less linked to a single social class.' The second aspect is a 'content analysis' of 'what is said and done within the power network'. Social science methods seeking to add case study and qualitative colour to their network analysis could do worse than supplement traditional collaborative methods with multi-method investigative approaches alive to the new possibilities that digitalization and social media affords. Exploring such opportunities is in keeping with Savage and Burrows call for 'greater reflection on how sociologists can best relate to the proliferation of social data gathered by others, which we currently largely ignore'. The key drivers of big data - digitalistion and globalisation - are not in themselves hostile to sociology and while they can be harnessed by powerful interests, they also offer ways to lift sociology to new insights. However, we do need to be aware of the potential pitfalls of what Ruppert (in this volume) terms the new big data ecologies, where our research becomes reliant on and vulnerable to the decisions of others as 'the devices and data we work with are framed, managed, configured by various other actors and interests'.

While we navigate big data ecologies as best we can there is also the question of not just accessing the data of others, but of creating our own datasets (which does not protect us from the vicissitudes of big data ecologies, but allows us the possibility of some order of control over how data may be managed, ordered, formatted and published). This is not beyond the reach of sociology. Indeed the cost of creating big data sets from publicly available online data is not of a fundamentally different order to that involved in a large sample of in depth interviews or indeed a representative sample survey. Obviously the integration of a variety of different datasets with existing datasets like the census, electoral roll, telephone directory, trustees of charities and directors of companies increases the time, expense and technical difficulties. But the promise of increasing digitization, data standardisation, machine readable data formats and the progress of Optical Character Recognition software, together with software with semantic capabilities to enable Natural Language Processing, will allow

the application of all sorts of sociological techniques to big data sets. Though the challenges vary, this point is as relevant for the sociologist interested in the social patterning of communications, knowledge or ideas as it is for those interested in the geographical or class patterning of health behaviours or outcomes.

Savage and Burrows remark that we live in 'an era when journalism is retreating from detailed social investigation' (2007: 895) and that sociology can provide specialized skills or knowledge. We agree that sociology has a role here, yet journalism is increasingly availing itself of new techniques of data driven journalism as can be seen in the examples of the *New York Times* data unit[1] and the *Guardian*'s data blog. The collaboration with the LSE on 'Reading the Riots' is an important indicator of options for the future. Such collaborations that span the academy, journalism and civil society, particularly in North America, have a longer pedigree than is often recognized. We can take the examples of the work of the Center for Responsive Politics, Public Citizen and the Center for Public Integrity in the US which have been enabled by the existence of transparency legislation to compile and analyse lobbying and campaign finance data disclosed by law. This work usefully bridges the academy and journalism, and charts a path that social science can surely contribute to, in terms of original research, knowledge exchange and indeed the impact agenda.

Big data has become a big issue for the funders of social research. In the UK the ESRC has prioritised capacity and innovation in this field, funding networks aiming to transform administrative data into evidence to inform policy making. Related to this is an initiative to similarly link and exploit data at a local level for businesses, public authorities, and the voluntary sector, and investigate rapidly evolving consumer behaviours and the social media sphere. Other funding agencies are also investing in big data related initiatives, including cyber and digital themes, such as the EPSRCs 'centres of excellence' in cybersecurity research. It is nevertheless clear that the social sciences and humanities are currently only in the foothills of the long ascent to the big data peaks. As scientific research funders promote interdisciplinary work we can anticipate, and perhaps celebrate, that sociology will be brought into conversation and collaboration with other disciplines. Investing our time and energy and securing funding for such an endeavour is no simple matter, but it is surely plain that it will be necessary to make strong and consistent arguments to the Research Councils and others on the importance of properly funding the gathering, processing and retention of big data. We perhaps need to stop obsessing with the defence of the discipline and drawing boundaries between subject areas, and seize the undoubted interdisciplinary possibilities that working with big data presents.

Big data is a social fact as well as a potential research approach and there is an urgent need to examine big data regimes directly as research objects. This might include reflexive work on the emerging trend towards collaborative big data projects (BBC 2013; Ansolabehere and Hersh 2012; ESRC 2013). Making sense of this explosion of information and data may be the impetus for new collaborations across higher education, and between academic and non-academic partners. Many of the big charitable foundations are seeking to understand and use ICT and social media to advance their programmes. They too are investing in projects that create and/or repurpose big data. With funding streams and data streams converging it is possible for the social sciences and humanities to lay claim to some of this new field. According to boyd and Crawford (2012) big data potentially offers these disciplines new ways to claim 'the status of quantitative science and objective method. It makes many more social spaces quantifiable. In reality working with Big Data is still subjective, and what it quantifies does not necessarily have closer claim on objective truth…Big data risks re-

inscribing established divisions in the long running debates about scientific method' (667) For us a sociological lens is highly relevant when it comes to the interpretation of big data. Our conceptual tools and theoretical insights represent the best available guarantors of our continuing relevance. However, we need to understand how to bring these to bear on the rapidly emerging possibilities in digital research and big data projects.

Investigative research, digital methods and big data
We can be optimistic about the possibilities for using big data to address enduring sociological issues (Webber 2009) and for social scientists to merge physical and virtual ethnographies (Murthy 2008; Murthy, this volume) integrating the internet and social media enabled methods. As Murthy observes 'a balanced combination of physical and digital ethnography not only gives researchers a larger and more exciting array of methods to tell social stories, but also enables them to demarginalize the voice of respondents in these accounts' (2008: 389). This is an enthusiastic assessment of the potential to engage with digital data. We certainly concur on the need to combine the virtual and real in research designs, and to carefully consider how to integrate different forms of data. Social scientists still need to follow Molotch's (1994) advice and go out into the field. They should also bear in mind his plea to ensure we have relevant and substantively worthwhile research questions to ask (what Murthy in this volume, drawing on Mills, calls the 'third camp' approach). It is likely that if sociology retains its relevance, concerns about the digital deluge will not spell the end for empirical social science. However, digital data and virtual ethnography need not only refer to the marginalised and powerless, and that one attraction of big data is that it offers new tools to research the powerful. While neither the interview nor the survey are yet passé, it is obvious that they are but two tools available to social scientists. Consider for instance how the analysis of texts and social networks can be reinvigorated using new digital tools to make sense of data found online or in digital databases.

We wholeheartedly endorse the interest of Savage and Burrows (2007: 895-6) in the 'politics of method', which we agree should involve sociologists 'renewing their interests in methodological innovation, and reporting critically on new digitalizations.' It is precisely in this area that we think that sociology has much to offer, including perspectives and research practices otherwise unavailable. Greater involvement in thinking about, using and reflecting on the politics of digital methods have the potential to open up new sociological fronts and to help to both understand and hold to account the producers and users of big data in the corporations and the state (see also Ruppert in this volume for an interesting discussion of the ethics and politics of big data).

The advent of big data, brings new methodological possibilities as well as challenges. In our view we are witnessing a change in both how the world operates and a consequent – required – change in the methods we use to understand it. We would point to the actual and potential role that sociology (and indeed other social sciences) can play in new developments in the analysis of language and meaning, as well as in social network analysis. Both of these techniques depend on well-established methodologies, though each has also taken on a new salience with the advent of digitization. While we do not seek to dismiss concerns about who owns and controls access to big data, it would equally be wrong to underplay some of the opportunities, especially those data sources which are freely and publicly available. We are already witnessing a growing interest in using data from social media (such as Facebook, Twitter, Instagram, Flickr, Pinterest, Reddit and the rest) and understanding what this tells us about new forms of action, social identity, solidarity, and community. While such data sources reveal aspects of popular culture and the dynamics of mass society, we also see the

possibilities of using the available sources to 'study up' (Williams 1989, Miller et al 2015) and conduct research on elites.

There is a growing literature on using 'social' and 'new' media to understand political processes, some of which focuses on elites. However, it is notable that much of this work is almost exclusively 'new media'-centric, examining how elected representatives use social media (Grant et al 2010), as part of permanent campaigning (Larsson 2014), the use of social media by election candidates (Graham et al 2014) government agencies (Waters and Williams 2011) or how interest groups use websites and Facebook (Nitschke et al 2014). What is largely missing in this literature is connecting social media data with other forms of evidence to explore elite dynamics. This seems to us the next phase in developing these lines of inquiry. The work of Burrows (including the chapter in this volume) on geodemographics and using Big Data to study the socio-spatialisation of High Net Worth Individuals (HNWI) is one means to study up. While the focus in that work is upon the super-wealthy we address political and corporate elites. There may be possibilities to bring together these streams of work to map the social, political and business networks of the powerful using big data.

One category of publicly available big data is the information routinely published and disclosed by governments, and data gathered using statutory mechanisms such as access to documents and freedom of information legislation. This provides a virtual treasure trove of data for social scientists studying elites, governance and policy-making processes. It appears to be a relative rarity for social scientists to request information from covered public agencies under the terms of Freedom of Information legislation. We tend to politely seek information from key gatekeepers in public office rather than insist on disclosure as a statutory right. Yet beyond the academy there has been a growth in 'monitory democracy', (Keane 2009) with civil society watchdog groups pooling, assembling and distributing information secured via FoI, or scraped from official data sources.[2]

One of the key drivers of this form of data is the implementation of various transparency regimes and mechanisms, in the service of good governance and democratic accountability. To illustrate this we would point to the impact of transparency measures in relation to lobbying and interest representation. The US has operated a lobbying disclosure system for several decades where lobbyists are required to submit quite detailed information on their staff, budgets, legislative interests and contacts. In the US there are also requirements to declare campaign finance contributions. Synthesising these data sources allows social scientists (as well as journalists, or indeed concerned citizens) to piece together how influence is brought to bear on democratic processes and decision-making. Dedicated civil society organisations have sifted and repackaged this information to make it intelligible for the non-expert. The Centre for Responsive Politics publishes the website www.openscrets.org which aggregates this data in the US, and offers a model that could be replicated in other countries. The trend towards transparency has meant that many other legislatures are implementing lobbying disclosure systems. Those interested in understanding the dynamics of power can begin to use this information – along with conventional social scientific research methods - to make sense of corporate strategy, and how companies seek to shape legislation and regulation. Equally, the policy activism of civil society can in part be traced using these sources. Bringing these together gives social scientists an empirically rich picture of the contest between the social movements for capital, ecology, and human rights over processes of globalisation.

Another promising source of data for researching the powerful is the disclosure of documents via litigation. Ever since the celebrated McLibel court case in the UK (http://www.mcspotlight.org/case/), in which London Greenpeace (not the global environmental group of the same name) was accused of libel by McDonald's, corporations have been careful to avoid reputational damage via disclosed documents. While some, like Shell in the case of the killing of Ogoni activist Ken Saro-Wiwi, have opted to settle out of court to avoid damaging publicity, in other cases it is individuals, consumers and victims who are taking the powerful to court, and in the process making internal documents public. Perhaps the most celebrated example of this is the tobacco archive (http://www.tobaccoarchives.com/), where as part of a successful class action brought against the major tobacco manufactures in the US, decades of internal industry documents were forced into the public domain. The archive contains over 26 million images of industry documents, and a further 1 million documents from the British American Tobacco archive in the UK have been integrated into the Legacy Tobacco Documents Library (LTDL). Notwithstanding the very significant resources required to enable public access to this information social scientists can now add to their knowledge of this industry based on internal, often highly sensitive, information that would not have otherwise been released to researchers.[3] The picture that emerges of the coordinated lobbying and public relations campaigns of the tobacco manufacturers to promote their products with politicians and consumers is not simply of historical interest, but sensitises researchers to contemporary corporate strategy. The Tobacco Tactics (http://www.tobaccotactics.org/) website is a good example of how such data can be analysed and condensed by academic researchers to help understanding of how corporations seek to manufacture doubt and counter public health. Disclosure via the courts is not confined to tobacco. Pharmaceutical companies have been compelled to release drug trial data and internal documents in a number of high profile cases, as catalogued by the Drug Industry Document Archive (http://dida.library.ucsf.edu/lawsuits.jsp) and there is a campaign to make the release of some information (though not raw data) on Randomised Control Trials mandatory (http://www.alltrials.net/).

Data Driven? Disclosure, transparency and digging deeper
There are a few converging key trends that are driving production and publication of new data that assists those seeking to study the powerful. The first is political, and relates to what is fast becoming an article of faith that good governance must be based on transparency, which facilitates democratic accountability as information is disseminated about decision-making processes and outcomes, and these information disclosures can be double-checked and analysed by social scientists, journalists, campaigning watchdog groups or others. This is a widespread and increasingly popular (at least among sections of officialdom and civil society) policy response to the acute crisis of trust and legitimacy that stalks (neo)liberal democracies. (Miller 2015) As the public grows increasingly sceptical about the good faith of political and business elites they are to be reassured that there is 'nothing to hide' through the routine disclosure of information and data about aspects of policy making, and the evidence base upon which discrete policies rest. Such disclosures are, of course, always contingent and mediated (Hansen and Flyverbom 2014), often partial, and rarely extend to private sector organisations in any meaningful way. Commercial confidentiality is often used as an exemption that precludes proper scrutiny of private entities. A parallel trend is politico-economic, and relates to the growth of e-government and reducing costs by delivering aspects of public services using new information and communication technologies (e.g. NHS 24). This trend in itself creates a wealth of information in digital format, which is potentially accessible via FoI, or may be shared with academics involved in evaluating performance and

delivery. Finally there is the social and technical, and the creativity that emerges from activists embracing new ICTs to source, scrape, sort and sift information. The burgeoning of this kind of data activism, combining elements of journalism, advocacy, crowd-sourcing, design, visualisation and mobilisation is the very essence of online campaigning to promote or address various political, social and environmental concerns. The pace at which these kinds of initiatives have developed, spread and adapted in line with technological advances over the last decade is quite remarkable and shows little sign of abating. If anything, it is likely to be a growth area as campaigning, data journalism, investigative reporting and social research cross-fertilise. It is nevertheless important to recognise that the growth of these forms of disclosure also have the potential to create an 'anticipatory transparency' entirely 'concerned with making predictions based on data mining and pattern recognition in large amounts of digital traces' (Hansen and Flyverbom 2014: 3).

To illustrate how disparate big data sources can be brought together we can point to a number of projects which use social network analysis to examine power elites and explore their connections and interlocks that are part of the structures of contemporary power. The first example examined data on the national origins, career roles, business and educational background of more than 14,000 speakers at over 800 events organized by the World Economic Forum between 2003 and 2009. The data were gathered from CV's posted on the WEF website in 2009 (and no longer available). (Schloegl et al 2013) This enabled a quantitative analysis of the background of the participants, showing, amongst other things the relatively narrow range of backgrounds (26% had attended one of 8 universities - all from the top 20 Universities as featured in global rankings). The most numerous contributors organisationally were from business with a pronounced emphasis on audit, financial and consultancy firms. The individuals that featured in the most meetings, however were predominantly from two sectors – media professionals and economics professors. This kind of data, freely – if sometimes temporarily - available and accessible to anyone with an internet connection and the right software, whether or not the 'owner' of the data wants it to be, can be used in sociological research precisely to examine global policy planning groups, comparing them with other groupings or with work on global corporate interlocks. (Heemskerk et al, 2013, Heemskerk 2013)

Such work focuses on the trans-nationality of European business elites and the networks clustered around the European Roundtable of Industrialists that constitute the 'hard core' of the global corporate elite (Carroll et al 2010). This gives us an empirically rich picture of elite business networks in Europe. The big data challenge is to link this information with databases like that on the WEF and others like the European Transparency Register (despite the very real short comings of this voluntary disclosure system, see ALTER EU 2013, 2015) or a range of other publicly accessible data. Both quantitative data like this and qualitative data generated through multi-method investigative research enable us to understand the political activities and strategies of the most powerful corporations and their directors. Such work could incorporate an analysis of the policy activism of the leading transnational corporations in Europe to produce a rounded account of corporate political networks and influence.

We began this paper recommending that sociologists dig deeper in their research. In this case we could suggest that using well-targeted freedom of information and access to documents requests may allow researchers to begin to examine in detail particular policy fields where parts of broader corporate networks closely engage with government. We can also resort to standard fieldwork methods, and seek to interview relevant actors and analyse documents in the public domain. Integrating quantitative and qualitative data sources – using secondary

data from public registers, disclosure systems, information on the web, twitter and networking sites like Linkedin, and primary data such as interviews and observational research at conferences and other policy relevant fora – means that there are many avenues available to study up and examine elite action in public policy. The published policy positions of elite organisations and specific corporations, along with the tweets and public statements of CEOs, corporate spokespeople, elite networkers and thought leaders, can give us new insights into elite ethos and ideology. Designing elite studies or policy work to include these kinds of different data sources will not only create big data archives, but will also offer us richer and deeper understandings of how policy and decision making are influenced. This will furnish useful knowledge that may allows us to better hold the powerful to account.

Conclusions

Whether leaked, scraped or officially disclosed, there are many exciting possibilities for new forms of elite studies using big data to facilitate sociological 'studying up'. One of the real possibilities opened up by publicly available data is to renew the tradition of power structure research on political and corporate elites. It appears to us that one of the main advantages of this form of digital research is that datasets can now be assembled without permission or co-operation, thus bypassing traditional difficulties in accessing elites.

While we are optimistic, we are aware that the era of big data presents considerable challenges for social scientists. The resource challenge is very real, and we need to strongly reject the myth that digital data on the web is somehow free to access and easy to manage. The costs involved in creating and maintaining big digital archives are significant, and well established monitoring projects like Open Secrets require an annual seven figure budget to continue their work. There is also a skills challenge that the discipline must respond to (see Murthy in this volume for a discussion of this topic). The technical know-how to scrape and clean data should not remain the preserve of computing experts. There is an urgent need to mainstream these techniques and expand the space for big data, social network analysis and investigative research on the curricula and methods courses designed by and for social scientists. This also suggests that we should work closely with disciplines that have something to teach us about scraping, cleaning, managing and manipulating digital data.

The ethics of digital research also need to be kept under review. Questions of informed consent and privacy have been at the centre of concerns about online vulnerability. With extensive terms and conditions attached to various online activities like participation on social media platforms the sovereignty of the individual with respect to their communications, opinions, and rights to association needs to be rethought. Some advocate an 'agile ethics' for big data research, where researchers are encouraged to claim responsibility for open datasets they have assembled (Neuhaus and Webmoor 2012). Power is clearly a factor in these debates: the power of corporations to use information that members of the public may think of as their own; the power of individuals to assert their right to be forgotten; and the power of researchers to secure access to data published online. Uldan (2014) shows how corporations have sought to monitor their critics online activities, illustrating how new forms of surveillance are available to those with the resources and interest to utilise them. In an unusual but revealing reversal, data for Uldan's research was obtained from the oil companies via subject access requests filed under the UK's Data Protection Act.

We suggest that there are emergent possibilities facing the social sciences and humanities to utilise big data, make it an object of study in its own right and embrace methodological

renewal and innovation. As big data already exists in the form of discovered, disclosed or digitised public repositories, our take on the challenge facing empirical sociology is to use this resource to critically investigate the powerful. This is 'decisive data' (Ruppert, this volume) in the sense that we can make choices in terms of our research questions and research designs that elevate or foreground particular social realities. Applying this to the challenge of studying up this means we can explore the opportunities presented by big data to better understand state, political and corporate elites and the governance processes they are embedded within and help constitute. To pursue such inquiry means we should be practically and temperamentally prepared to dig deeper. The purpose of such work is neither simply academic nor activist, but should also have practical relevance in that it assists accountability. We can learn from and with others who are using investigative approaches to scrutinise the powerful. There are clear possibilities here to practice public sociology (Burawoy 2005), engaging with wider civil society organisations and movements, many of which are coalescing around concerns over privacy, surveillance, transparency and accountability. We think this provides one answer to the key question of relevance facing sociology in an era where there is ever greater private encroachment on higher education.

References

ALTER EU (2013) 'Rescue the Register! How to make EU lobby transparency credible and reliable', Brussels, 20 June. http://www.alter-eu.org/documents/2013/06/rescue-the-register (accessed 10 October 2014).

ALTER EU (2015) 'New and Improved? Why the EU Lobby register still fails to deliver' Brussels 28 January. (accessed 29 January 2015)

Ansolabehere S and Hersh E (2012) Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate *Political Analysis* (20):437–459.

BBC (2013) The Great British Class Calculator, *BBC News Magazine*. 3 April http://www.bbc.co.uk/news/magazine-22000973

boyd D and Crawford K (2012) 'Critical Questions for Big Data: Provocations for a cultural, technological and scholarly phenomena'. *Information, Communication and Society* 15(5): 662 – 679.

Burawoy M (2005), '2004 American Sociological Association Presidential address: For public sociology', *The British Journal of Sociology*, 56 (June): 259–294.

Carroll W, Fennema M and Heemskerk E (2010) 'Constituting corporate europe: A study of elite social organization'. *Antipode* 42: 811–843.

Centre for Responsive Politics, (nd) *Opensecrets* http://www.opensecrets.org/lobby/ Washington DC.

Diamond S (1992) *Compromised Campus: the collaboration of universities with the intelligence community, 1945-55*, Oxford: Oxford University Press.

Domhoff WG (2005) *Who Rules America? Power, Politics, and Social Change*, 5th edition, New York: McGraw-Hill.

ESRC (2013) Researchers team up with Google to explore data', http://www.esrc.ac.uk/news-and-events/features-casestudies/features/25707/Researchers_team_up_with_Google_to_explore_data.aspx

Graham T, Jackson D and Broersma M (2014) 'New platform, old habits? Candidates' use of Twitter during the 2010 British and Dutch general election campaigns'. *New Media & Society*.

Grant W, Moon B and Grant J (2010) 'Digital dialogue? Australian politicians' use of the social network tool Twitter'. *Australian Journal of Political Science* 45(4): 579–604.

Hansen HK and Flyverbom M. (2014) 'The politics of transparency and the calibration of knowledge in the digital age.' *Organization*. 00 (accessed 10 October 2014).

Hayes M and McIntyre A (2014) 'The Liverpool University Research Project on "Dissident" Irish Republicanism – Working for the British State?' *The Pensive Quill*, October. https://web.archive.org/web/20141012101057/http://thepensivequill.am/2014/10/the-liverpool-university-research.html Retrieved from the Internet Archive of 12 October 2014 (accessed 12 October 2014).

Heemskerk EM (2013) 'The rise of the European corporate elite: evidence from the network of interlocking directorates in 2005 and 2010'. *Economy and Society* 42: 74–101.

Heemskerk EM, Daolio F and Tomassini M (2013) 'The Community Structure of the European Network of Interlocking Directorates 2005–2010'. *PLoS ONE* 8(7).

Herman E (1998) 'Project Camelot and the career of cold war psychology' in Simpson, C. (ed) *Universities and Empire: money and politics in the social sciences during the Cold War*, New York: New York Press.

Keane J (2009) *The Life and Death of Democracy*, London: Simon & Schuster.

Larsson AO (2014) 'Online, all the time? A quantitative assessment of the permanent campaign on Facebook'. *New Media & Society*.

McKie L and Ryan L (2012) 'Introduction to E-Special: Exploring Trends and Challenges in Sociological Research', *Sociology*, E-Special Issue 1: 1-7.

Miller D (2013) 'Spying on academics will not help fight terrorism', *The Conversation*, 15 July 2013, 6.42am BST. https://theconversation.com/spying-on-academics-will-not-help-fight-terrorism-16046

Miller D (2015) 'Neoliberalism, Politics and Institutional Corruption: Against the "Institutional Malaise" Hypothesis', in David Whyte (ed) *How corrupt is Britain?*, London: Pluto Press.

Miller D and Mills T (2010) 'Counterinsurgency and terror expertise: the integration of social scientists into the war effort' *Cambridge Review of International Affairs*, Volume 23, Number 2, June: 204-221.

Miller D, Mills T and Harkins S (2011) 'Teaching about terrorism in the UK: how it is done and what problems it causes', *Critical Studies on Terrorism*. 4(3): 405-20.

Miller D, Brown L, Dinan W and Stavinoha L (eds) (forthcoming) *Researching the Powerful: Public Sociology in Action*, London: Routledge

Molotch H (1994) 'Going Out', *Sociological Forum,* 9(2): 221-239.

Murthy D (2008) 'Digital Ethnography: An Examination of the Use of New Technologies for Social Research', *Sociology*, 42(5): 837–855.

Nader, L. (1972) 'Up the Anthropologist: Perspectives Gained from 'studying up'' in D. Hymes (ed) Reinventing Anthropology New York: Random House: 284–311.

Neuhaus F and Webmoor T (2012) 'Agile ethics for massified research and visualisation'. *Information, Communication and Society* 15(1): 43–65.

Nitschke P, Donges P, and Schade, H (2014) 'Political organizations' use of websites and Facebook', *New Media & Society*. August.

Savage, M. & Burrows, R. (2007) 'The Coming Crisis of Empirical Sociology', *Sociology* Volume 41(5): 885–99.

Savage M and Burrows R (2009) 'Some further reflections on the coming crisis of empirical sociology', *Sociology* 43(4): 762–72.

Schloegl M, Plehwe D and Miller D (2013) 'The intellectuals of capital: Towards a political sociology of the World Economic Forum?', Paper presented at Societies in Conflict: Experts, Publics and Democracy, the 8th International Interpretive Policy Analysis Conference, University of Vienna, Austria, 4 July.

Simpson C (1994) *Science of Coercion: Communication Research and Psychological Warfare 1945-1960*. New York: Oxford University Press.

Stapleton, L. K. (2011). 'Taming Big Data'. *IBM Data Management Magazine*, 16(2):12-18,

Uldam J (2014) 'Corporate management of visibility and the fantasy of the post-political: Social media and surveillance.' *New Media & Society*, 00: (accessed 10 October 2014). http://nms.sagepub.com/content/early/2014/06/30/1461444814541526

Waters R and Williams J (2011) 'Squawking, tweeting, cooing, and hooting: Analyzing the communication patterns of government agencies'. *Journal of Public Affairs* 11(4): 353–63.

Webber R (2009) 'Response to The Coming Crisis of Empirical Sociology: An Outline of the Research Potential of Administrative and Transactional Data' *Sociology* 43(1): 169-78.

Williams K (1989) 'Researching the powerful: problems and possibilities of social research'. *Contemporary Crises*, 13(3), 252-274.

1**Endnotes**

 See Open: All the Code That's Fit to Print, *New York Times*, http://open.blogs.nytimes.com/

2 See, for instance, on the UK: They Work For You (http://www.theyworkforyou.com/), What Do They Know (https://www.whatdotheyknow.com/), on the EU: Farmsubsidy.org (http://farmsubsidy.openspending.org/), OpenInterests.eu (http://openinterests.eu/), Ask the EU (http://www.asktheeu.org/); and on the US: Open Secrets (https://www.opensecrets.org/); the FOIA Machine (https://www.foiamachine.org/)

3 For an overview of the effort to make the Guildford repository accessible see British American Tobacco Documents Archive, Project History. http://bat.library.ucsf.edu/history.html.